ESD-TR-91-267

MTR-11194

Performance Parameter Methodology Update for Information Delivery Times

By

B. H. Roberts

March 1992

Prepared for

Deputy Director, Systems Engineering
Space & Missile Warning Systems Program Office
Electronic Systems Division
Air Force Systems Command
United States Air Force
Hanscom Air Force Base, Massachusetts

Project No. 4370

Prepared by

The MITRE Corporation
Bedford, Massachusetts

Contract No. F19628-89-C-0001

REVIEW AND APPROVAL

This technical report has been reviewed and is approved for publication.

Gerard C Cavallo

GERARD CAVALLO, CAPT, USAF
Systems Engineer,
Space and Missile Warning
Program Office

FOR THE COMMANDER

Timothy R Moulton

TIMOTHY R. MOULTON, GM-15
Deputy Director, Systems Engineering
Space and Missile Warning
Systems Program Office

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
March 1992

3. REPORT TYPE AND DATES COVERED
Final

4. TITLE AND SUBTITLE
Performance Parameter Methodology Update for Information Delivery Times

5. FUNDING NUMBERS
F19628-89-C-0001
4370

6. AUTHOR(S)
Roberts, Barbara H.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The MITRE Corporation
Burlington Road
Bedford, MA 01730

8. PERFORMING ORGANIZATION REPORT NUMBER
MTR-11194

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Deputy Director, Systems Engineering Space & Missile Warning Systems Program Office (ESD/SR-1)
Electronic Systems Division, AFSC
Hanscom AFB, MA 01731-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ESD-TR-91-267

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
This report documents the development of a methodology for determining the overall information delivery time specifications for the Cheyenne Mountain Upgrade (CMU) from the individual subsystem transit time specifications. Since little is known about the transit time distribution of each subsystem, each is approximated by a rectangular distribution. Then these rectangular distributions are convolved to determine the distribution of the overall system performance parameter, information delivery time. When many rectangular distributions must be convolved, a simplification employing the Central Limit Theorem is used to estimate the mean and variance of the system's overall distribution. This distribution is thought to be close to normal with no negative tail. When rectangular distributions with large variance are used, errors in the resulting distribution can occur. These errors, introduced by the simplification technique, are reduced either by using numerical methods for convolving discrete distributions sampled from the continuous curve or, when less accuracy is required, by an approximation technique involving the convolution of two uniform distributions. A comparison of the numerical results derived from this approach with results calculated in a previous study is also presented.

14. SUBJECT TERMS
Information Delivery Time
Probability Density Function
System Performance

15. NUMBER OF PAGES
63

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

ACKNOWLEDGEMENTS

This document has been prepared by The MITRE Corporation under Project No. 4370, Contract No. F19628-89-C-0001. The contract is sponsored by the Electronic Systems Division, Air Force Systems Command, United States Air Force, Hanscom Air Force Base, Massachusetts 01731-5000.

The author extends her thanks to Richard Bentz (D052), Darwyn Harris (D045), Dr. Lee Parker (D085), Dr. Robert S. Swarz (D052), Dr. Alan Vogel (D045) and Dr. David Yin (D096). Special thanks are extended by the author to John Kriegel (D085) and Dr. James W. Matthews (D050) for their contributions through extensive technical and collegial conversations. Major programmatic and system related contributions, for which the author is appreciative, were made by Ernest. I. Perelmuter (D041) and Arno E. Schiegoleit (D041).

TABLE OF CONTENTS

SECTION **PAGE**

LIST OF ILLUSTRATIONS

FIGURE		PAGE

LIST OF TABLES

SECTION 1

INTRODUCTION

1.1 BACKGROUND

The five Cheyenne Mountain Upgrade (CMU) acquisitions were each specified independently with no end-to-end performance requirements. In May 1989, the CMU acquisitions were combined under a single Program Management Directive (PMD) for CMU programs. This action defined the CMU as a subsystem of the overall Integrated TW/AA system. In the summer of 1989, five key CMU-level performance parameters were derived for inclusion in the CMU Acquisition Program Baseline (APB), Decision Coordination Paper (DCP) and Test and Evaluation Master Plan (TEMP). This action was accomplished under ESD/AFSPACECOM coordination. The coordinated parameters were then adopted by AFSPACECOM for inclusion in the CMU System Operational Requirements Document (SORD).

Estimates for the five, key CMU-level performance parameter values are updated quarterly. One of the five key parameters is information delivery time. Questions relating to the methodology for estimating both the subsystem transit times and the overall information delivery times were raised and a request to review the methodology was made. This paper documents the updated information delivery time methodology.

1.2 OBJECTIVES

In the context of updating the CMU performance parameter methodology, the representation of each subsystem's message transit time and the derivation of the overall information delivery time from these representations are investigated. The goals of this study are to develop a methodology for deriving the overall CMU information delivery time specification from the transit time specifications of each of the subsystems, to reevaluate the assumption of the normal distribution for each of the subsystem transit times, to identify alternative distributions for the subsystem transit times, and to assess other related methodological considerations.

1.3 SCOPE

Transit time, depicted in figure 1, is defined as the time a message takes to transit a CMU subsystem. Information delivery time is the total time from message input to the CMU to corresponding CMU message output to the Forward User display systems. The suite of subsystems that contribute to the overall information delivery times differ for the three missions. These three missions are the Missile Warning (MW), the Air Warning (AW) and the Space Warning (SW) missions. The strings associated with each of these missions are shown in figure 2. The following methodology does not address transit times for the sensors nor communication propagation times. Message transmission times are treated as constants. Future work will address the variations in transmission times for the various links.

M. H. Weeden

IL2935/refVL14639

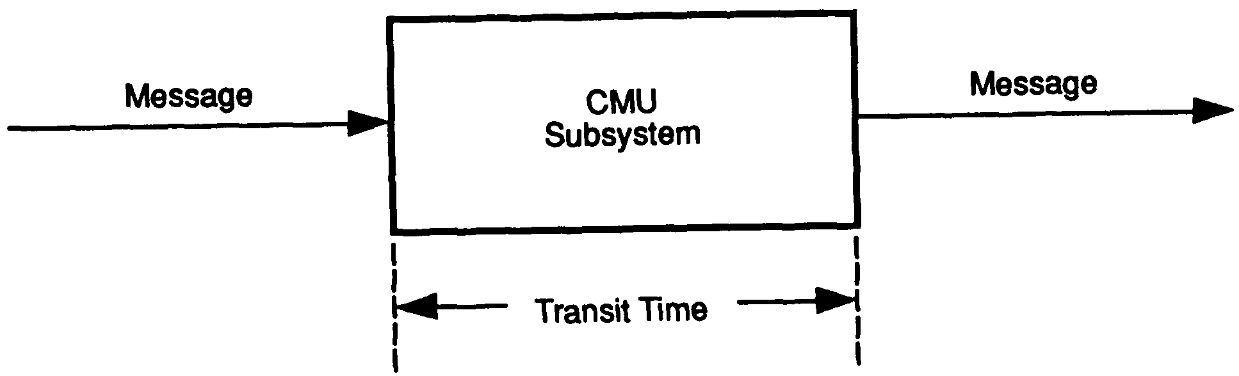


Figure 1. Transit Time

IL2936/refVL15035

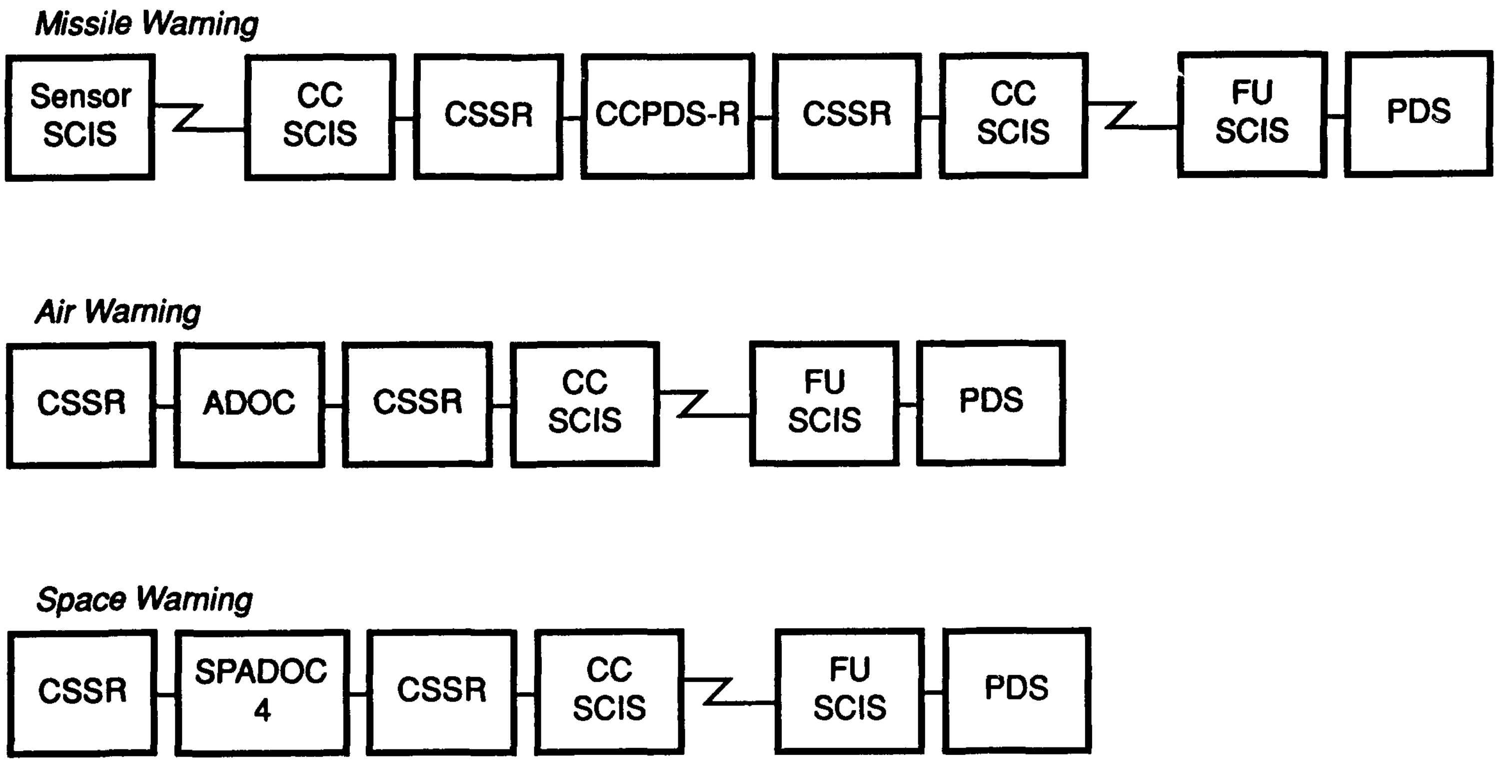


Figure 2. The Strings

1.4 APPROACH

Since each of the CMU subsystem specifications were developed independently, no statistical consistency in the transit time specifications exists. Furthermore, the requirements for missile, air and space warning information delivery time have not been specified. The approach recommended in this paper for specifying the overall information delivery time uses rectangular distributions to represent the subsystem specifications. Convolution of the rectangular distributions approximating the subsystem transit times is the preferred process for deriving the overall information delivery times. Since the convolution of a large number of distributions is difficult, the Central Limit Theorem is used to approximate the convolution of these rectangular distributions. The resultant distribution, derived from the convolution product, represents the overall information delivery time distribution.

The following sections address the overall information delivery time derivation methodology. An overview of the subsystem transit time specification can be found in section 2. Previous analysis is investigated along with the original assumptions used in that work in section 3. In section 4, an investigation of the SCIS test data is explored. In section 5, approximate solutions are explored and explained. Conclusions and recommendations follow in sections 6 and 7.

SECTION 2

THE OVERALL INFORMATION DELIVERY TIME SPECIFICATION PROBLEM

2.1 THE TRANSIT TIME SPECIFICATIONS

In each of the subsystems, the transit time is specified differently. CSSR specified transit time at the 99.8 percent and 99.99 percent levels. The SCIS program required the messages to transit the subsystem in "a" seconds 99 percent of the time with a maximum of "b" seconds. A total queue time was also specified. The SPADOC subsystem specified a mean and maximum transit time. The CCPDS-R program specified the 98 percent value. Granite Sentry specified a Space Warning transit time at 95 percent and an Air Warning Transit time at 95 percent. Since the specification values are classified, the real values of the specifications have been replaced with variables as place holders in figure 3.

2.2 LACK OF INFORMATION IN SUBSYSTEM SPECIFICATION

In three of the subsystems, the specification of the transit time is given as one percentile only, making it difficult for the analyst to determine the transit time distributions. The provision of one point is not adequate to determine the normal curve for that subsystem since the normal curve is completely determined by its mean and standard deviation. The determination of these two parameters by solving two simultaneous equations needs two percentile points. Without a second point, guesses have to be made about the normal distribution parameters. These guesses may significantly affect the resultant information delivery time distribution as shown in figures 4 and 5.

The guesses can have a significant, and inaccurate impact, especially for distributions that are clustered around the positive side of zero. In the past, when the assumption of a normal distribution was employed, negative transit times were sometimes found for reasonable percentiles. There was no computational error in these calculations. The error was in the assumptions. These negative transit time values arose when both a mean and a variance were assumed for the normal distribution. These spurious findings led to further investigations of the transit time methodology.

2.3 SPECIFICATION OF OVERALL INFORMATION DELIVERY TIME

The problem is to derive the overall information delivery time specification from these program specifications. A graphic depiction of the meaning of the overall information delivery time is given in figure 6. The difficulty associated with this task is that there is very little information for the analyst to use in performing this derivation.

In the next section the original approach is discussed, along with its basic assumptions and its methodology.

IL2976

SCIS

99% a seconds
100% b seconds
a queue

CSSR

99.8% c seconds
99.99% d seconds

CCPDS-R

98.5% e seconds

SPADOC

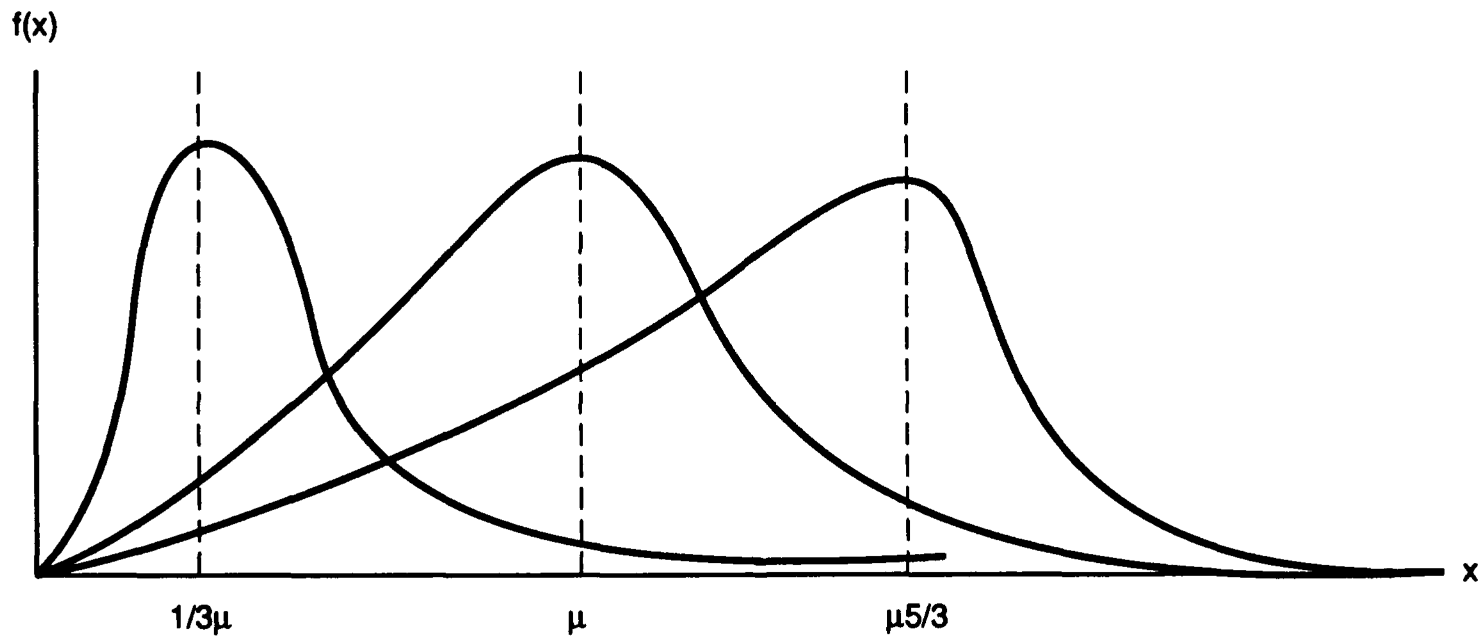
Mean f seconds
Max g seconds

GS II

95% h seconds

Figure 3. The Specification Values

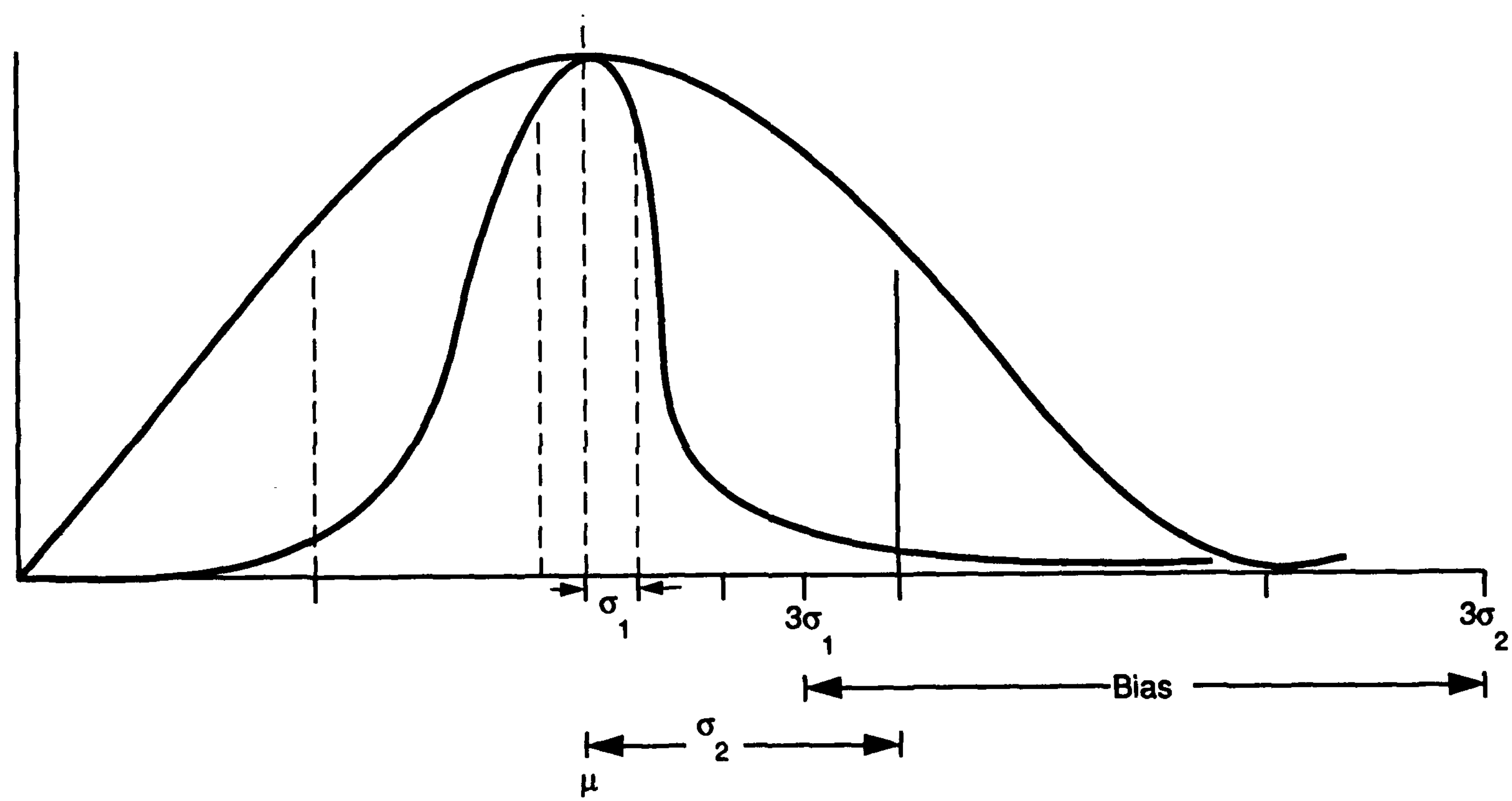
IL2937/refVL15012



- The choice of the mean has significant effect on the answer

Figure 4. Effect of Error in Estimating Mean

IL2938/refVL15013



● The choice of the variance has significant effect on the answer

Figure 5. Effect in Error in Estimating Variance

IL2962/refVL15143

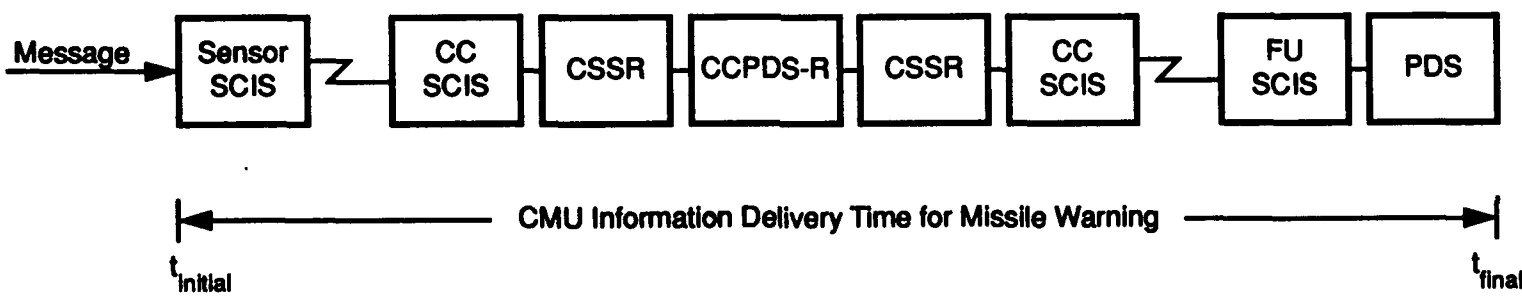


Figure 6. Overall Information Delivery Time for Missile Warning

SECTION 3

PREVIOUS STUDIES

3.1 THE ORIGINAL APPROACH

The current approach to determining the overall CMU information delivery time specification for the APB is to 1) assume that the transit time distribution for each subsystem is Gaussian or normal, 2) find the mean and variance for each of these distributions from the specification values, 3) normalize each of the subsystem transit times to 99.8 percent, and then 4) add these 99.8 percent values to get the overall information delivery time for a particular string.

3.2 ORIGINAL APPROACH ASSUMPTIONS

One of the disadvantages of the current approach is that the normal distribution has a negative tail. The existence of a negative portion of the distribution means that it is possible for messages to take negative time in transiting a subsystem. In addition, a number of the subsystems have small, specified transit times. Using the current approach, negative values for reasonable percentiles during normalization have been found. These findings are a clear indication that the assumption of the normal curve may be incorrect, at least for some of the subsystems.

A second shortcoming with the current approach is the addition of the 99.8 percent values of each of the individual message transit times for the subsystems in the particular string under consideration to get an overall 99.8 percent information delivery time. Mathematical theorems exist that say 1) the means of any distributions can be added to get an overall mean, and 2) if the subsystems are independent, the variances can be added to yield a variance for the overall distribution. Except for means, and in the case of independent subsystems, variances, the sum of random variables x and y should be found using convolution or an approximation to the convolution.

3.3 METHOD FOR ADDING RANDOM VARIABLES

The method for adding random variables is specified in *Probability, Random Variables, and Stochastic Processes* by Papoulis. [1] The Fundamental theorem states: "If the random variables x and y are independent, then the density of their sum $z = x + y$ equals the convolution of their respective densities." [1] By densities, Papoulis means distributions, or basically histograms. (A histogram is a graph with, in this case, transit times on the independent axis and counts of the occurrences of these times along the dependent axis.) The Fundamental Theorem means that the analyst can not add transit times at a specific percentile from different distributions and report the sum as either a worst case or the corresponding percentile.

3.4 CLARIFYING EXAMPLE

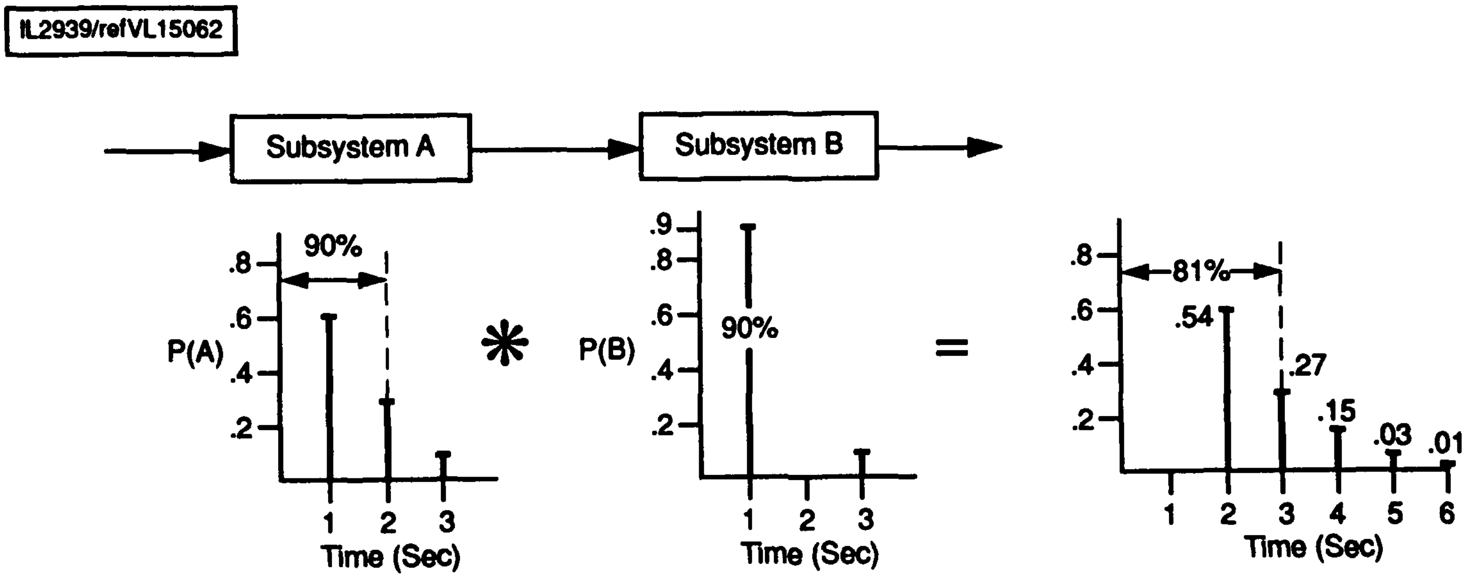
The analyst needs to convolve the probability density functions for the subsystems because of the nature of the statistical problem. To clarify, consider a simple problem. Suppose that there are two random variables, x and y. Suppose that their sum is z, i.e. z=x+y. Also suppose that z equals 10 and consider only the integers between 1 and 10. Then the sum, 10, can be acquired by the sum of either 1 and 9, or 2 and 8, or 3 and 7, and so on. This is the same as either 1 and (10-1), or 2 and (10-2), or 3 and (10-3), etc. Thus, the sum 10, can be arrived at through the sum of a number of different values. Since we are looking for the probability of the occurrence of 10, the sum of the product of the probabilities of the various summands is the probability of the occurrence of 10. The above algorithm is a simple example of the discrete convolution product. There are two exceptions for adding random variables. It can be shown that the means can be added to obtain the mean of the overall distribution [1]. In the case of independence, variances can also be added to get the variance of the overall distribution [1].

3.5 COUNTEREXAMPLE

In order to further examine this counterintuitive concept, consider the following counterexample depicted in figure 7. In that example, the addition of the 90 percent value in the first distribution and the 90 percent value in the second distribution gives the 81 percent in the overall distribution. In the figure, the system consists of two simple subsystems, A and B. In subsystem A, messages transit the subsystem in three discrete times. These times are 1, 2, and 3 seconds. A count of the number of messages that transited subsystem A in 1 second reveals that they occur with a probability of .6. (This means that if 100 messages were sent through subsystem A, that it is likely that 60 would get through in 1 second.) In subsystem B, 90 percent of the messages get through in 1 second, none get through in 2 seconds, and 100 percent get through in three seconds. Convolution of the two distributions produces the overall distributions on the right of the figure.

The values for the convolution are found at the bottom of the figure. For the overall transit time of 2 seconds, there is only one way to get a transit time of 2 seconds. This way is to transit the first subsystem in one second and the second subsystem in one second. The probabilities, .6 and .9, are multiplied to get .54. For a transit time of 3 seconds, either the message transited the first subsystem in 1 second and the second subsystem in 2 seconds or the first subsystem in 2 seconds and the second in 1 second. The probabilities for these are displayed and add to .27. This technique is continued for all possibilities.

Observe that even though 90 percent of the messages transit the first subsystem in 2 seconds and the second in 1 second, only 81 percent of the messages transit the overall system in 3 seconds, the sum of 1 and 2 seconds. This simple example shows that at least in some cases, convolution produces a different result than simply adding the 90 percent values.



Σ = A + B	P(A)	P(B)	P(A) • P(B)
2 = 1 + 1	.6	.9	.54
3 = 1 + 2	.6	0	.27
2 + 1	.3	.9	
4 = 1 + 3	.6	.1	.15
2 + 2	.3	0	
3 + 1	.1	.9	
5 = 2 + 3	.3	.1	.03
3 + 2	.1	0	
6 = 3 + 3	.1	.1	.01

Figure 7. Counterexample

3.6 ORIGINAL METHODOLOGICAL SUMMARY

Thus, the original methodology assumed that the transit time distributions were normal. Then, when necessary, means and variances for these distributions were assumed. The 99.8 percent defacto standard found by normalization techniques was calculated, and then these 99.8 percent values for each of the subsystems were added in the order in which messages traverse the system. This sum became the overall information delivery time. In the next sections, the updated methodology is presented.

SECTION 4

INVESTIGATION OF TEST DATA DISTRIBUTIONS

4.1 SAMPLING TECHNIQUE

Early in the methodology update analysis, samples of the SCIS transit time data were taken. The intent was to determine the transit time distribution(s) from SCIS test data. Five histograms of 100 message transit times were developed and compared to determine if there was significant variance between samples. The observed variation in samples was within reasonable limits.

4.2 HISTOGRAM DEVELOPMENT

A composite histogram of the 500 data points was compiled from the individual histograms and the mean and variance of the test data found. The resultant histogram, with some trimodal characteristics evident, is shown in figure 8. For the sample size of 500 test values, the corresponding Gaussian distribution is shown overlaid on the test data. (The CSSR project also provided the analyst with the mean and variance of CSSR test data for message transit times. These were used as a "zeroth order filter" to a candidate set of distributions as is explained below. Preliminary calculations indicate that the CSSR message transit times may be Erlang. A comparison of the possible CSSR Erlang distribution is also shown in the figure 9.)

4.3 CANDIDATE DISTRIBUTIONS

A list of possible distributions that might fit the SCIS test data histogram was postulated. These candidate distributions were 1) the Erlang, 2) the Exponential, 3) the displaced Exponential, 4) the Normal or Gaussian, 5) the Maxwell, 6) the Rayleigh, and 7) the Gamma function. These distributions are depicted in figures 10 through 16.

The Erlang was developed to model telephone situations and was deemed to be a good candidate. In the Erlang distributions, messages arrive at a service or processing facility where k distinct phases of the service must be performed on each customer. If the time to perform each phase has an exponential distribution and is independent of the time to perform each of the phases, then the total time to perform all k phases of service has the Erlang distribution.

The normal distribution was selected as a possibility because 1) it was currently being used and 2) if the mean were far enough to the right, it might not be a bad choice. It is well known that in situations where a lot of unknown variables are working on a parameter, that the Gaussian is often a good choice.

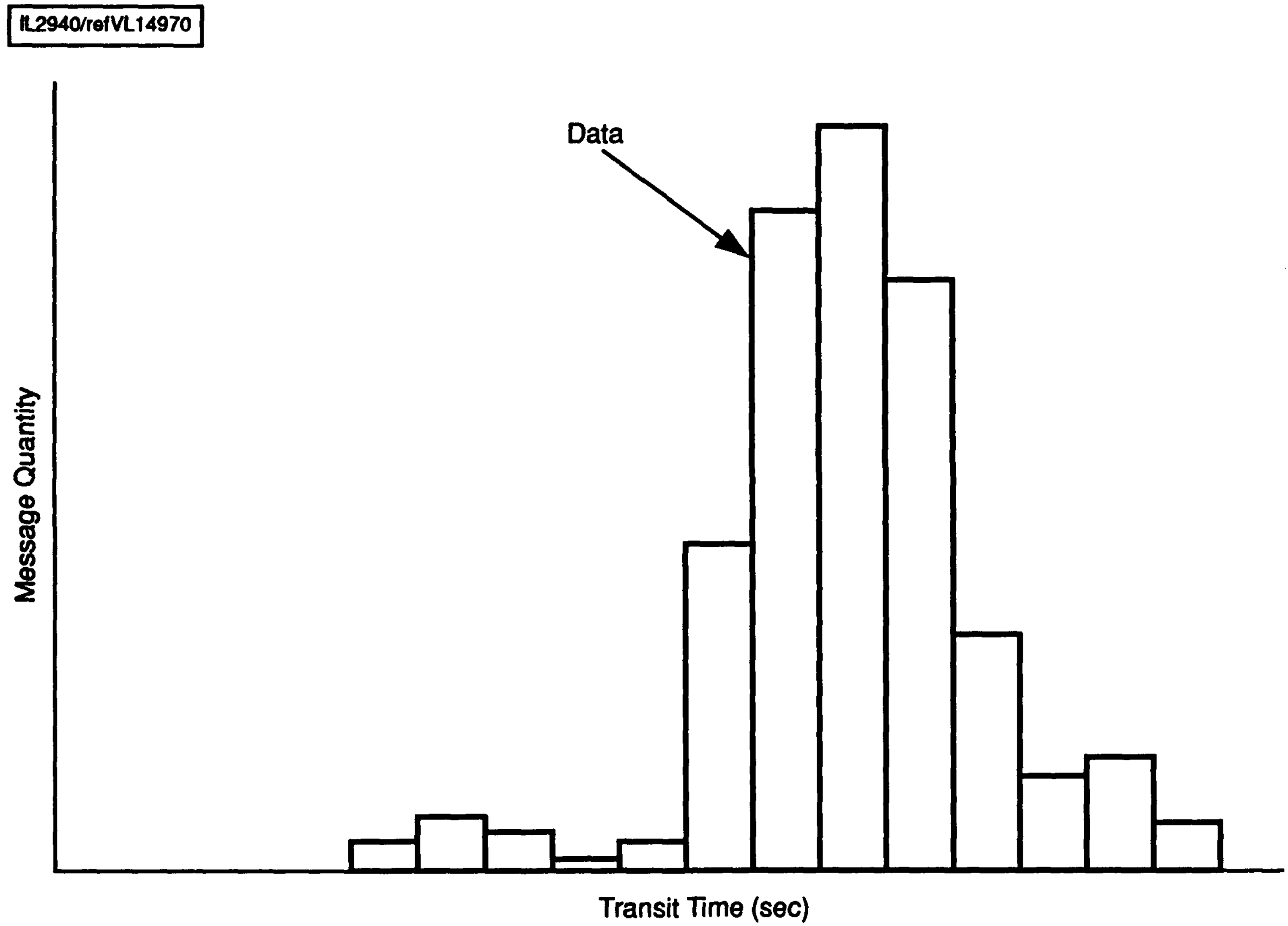


Figure 8. SCIS Test Data Histogram

IL2941/refVL14640

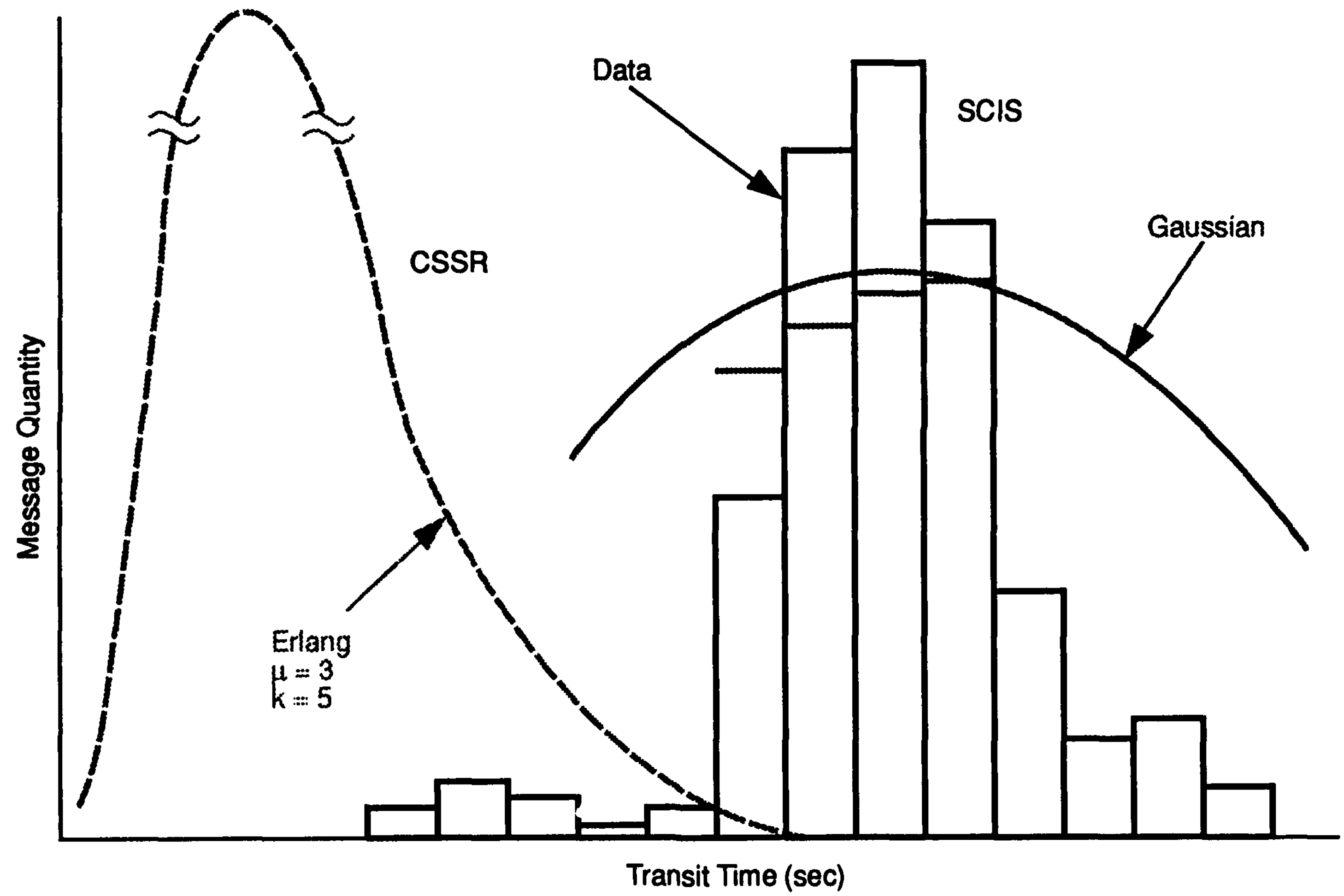
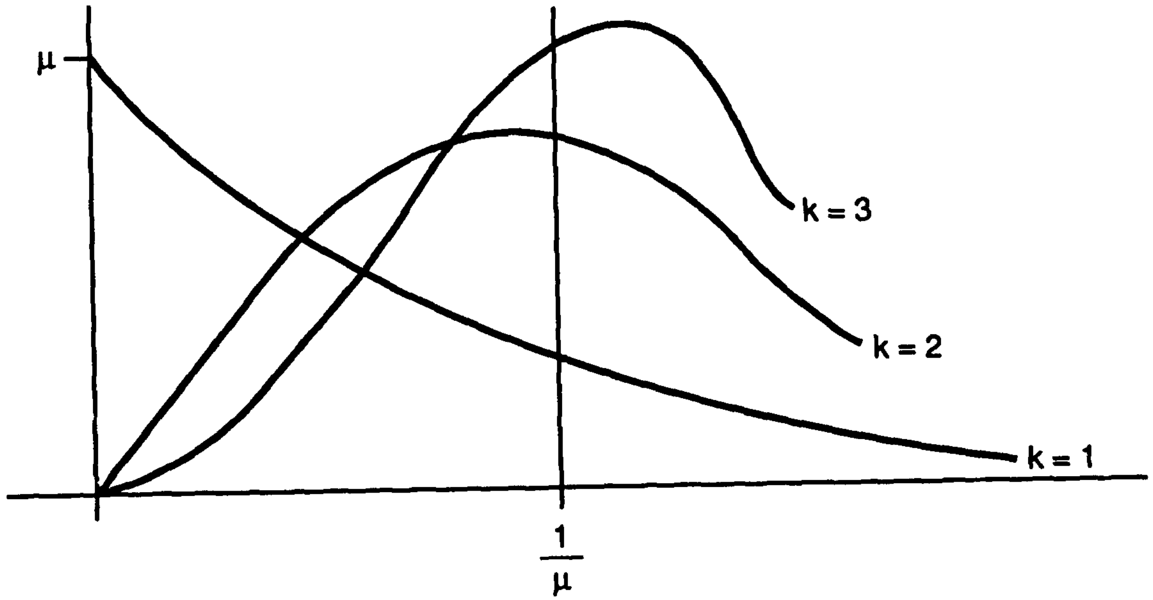


Figure 9. CSSR Test Data – Preliminary Results

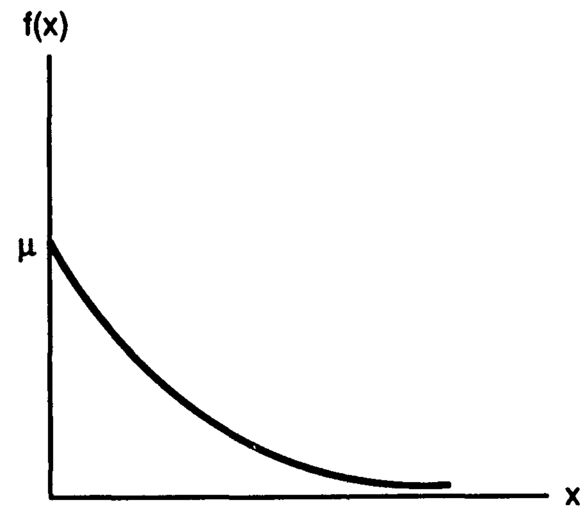
IL2942/refVL14404



Customers (message) arrive at a service facility (processing) where k distinct phases of the service must be performed on each customer. If the time to perform each phase has an exponential distribution $f(t_i) = k\mu e^{-k\mu t_i}$ and is independent of the time to perform each of the phases then the total time T to perform all k phases of service has the Erlang distribution

Figure 10. The Erlang Distribution

IL2943/refVL14643



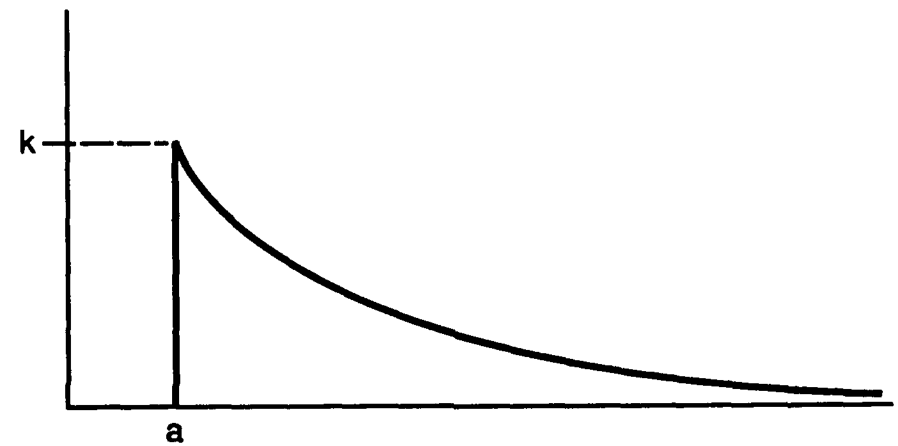
$$f(x) = \mu e^{-\mu x} \qquad x \geq 0,\ \mu \geq 0$$

$$E(X) = \frac{1}{\mu}$$

$$Var(X) = \frac{1}{\mu^2}$$

Figure 11. The Exponential Distribution

IL2944/refVL14702



$$f(x) = \mu e^{-(\mu + a)t}$$

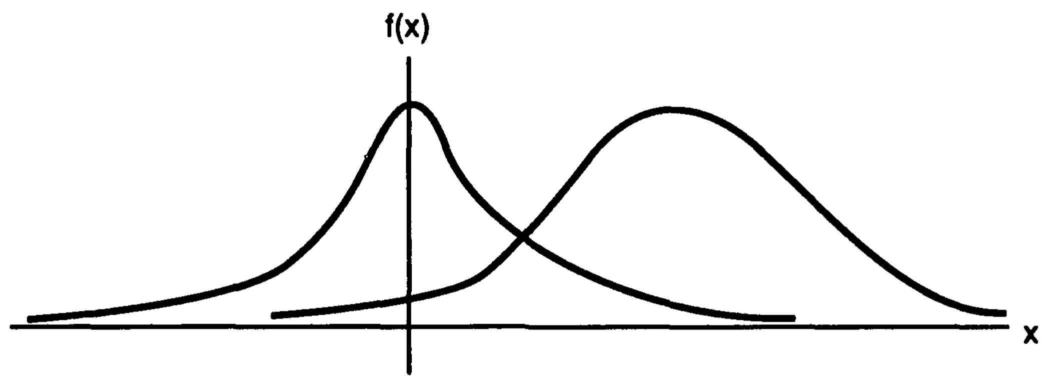
$$= \mu e^{-at} e^{-\mu t}$$

$$= k e^{-\mu t}$$

Figure 12. The Displaced Exponential

M. H. Weeden

fL2945/ref VL14641

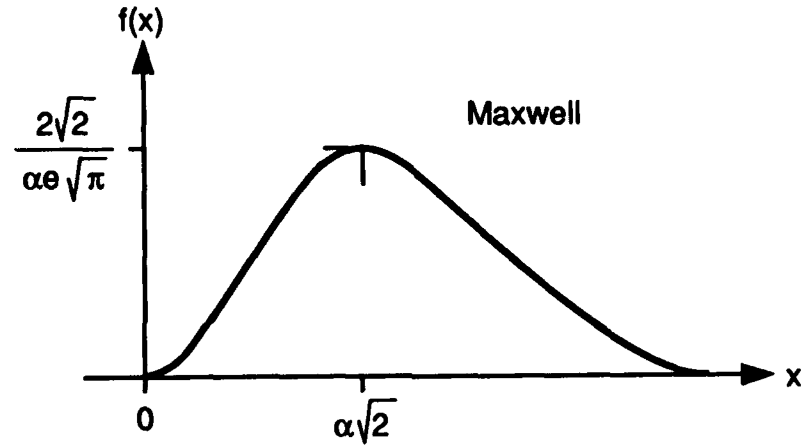


$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \; e^{-1/2\,[(x-\mu)/\sigma]^2}$$

- Advantages: Well-known, easy to use
- Disadvantages: Negative transit times

Figure 13. The Normal Distribution

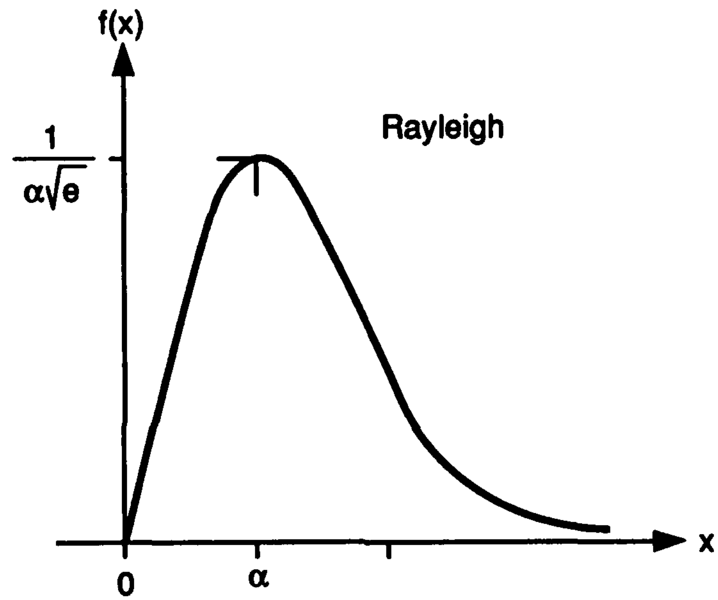
IL2946/ref VL14645



$$f(x) = \frac{\sqrt{2}}{\alpha^3\sqrt{\pi}}\, x^2 e^{-x^2/2\alpha^2}\, U(x)$$

Figure 14. The Maxwell Distribution

IL2947/ref VL14644

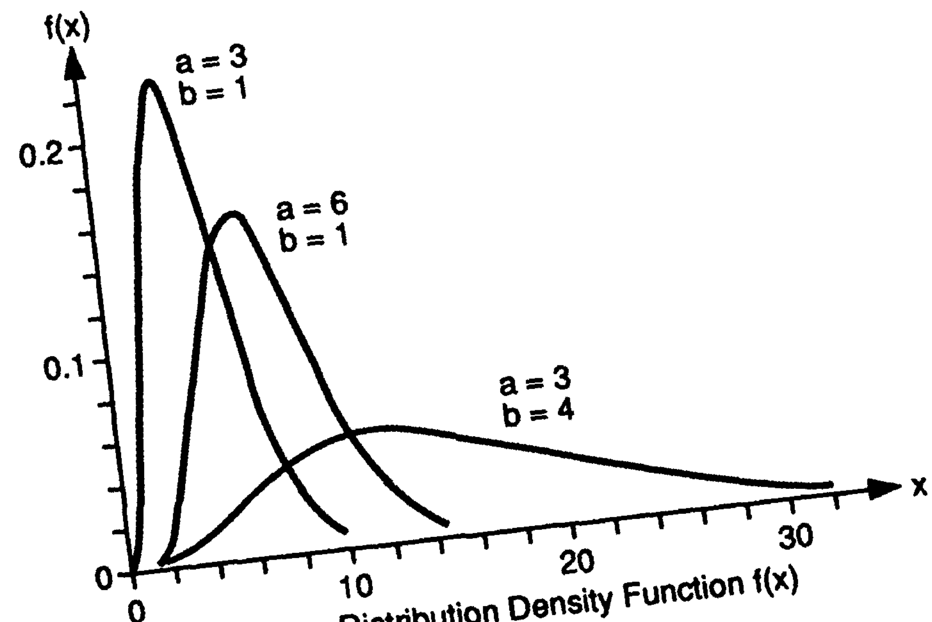


$$f(x) = \frac{x}{\alpha^2} e^{-x^2/2\alpha^2} U(x)$$

Figure 15. The Rayleigh Distribution

M. H. Weeden

IL2948/ref VL14642



Gamma Distribution Density Function f(x)

$f(x) = [b^{a+1}\,\Gamma(a+1)]^{-1}\,x^a e^{-x/b}$ $b > 0,\ a > -1$, for $x > 0$

$= 0$ for $x \le 0$

$E(X) = b(a + 1)$ $a = 50$

$Var(X) = b^2(a + 1)$ $b = 0.025$

- Advantage: General
- Disadvantage: Complex

Figure 16. The Gamma Function

The exponential and displaced exponential were chosen because these are the standard distributions used in queueing theory applications for the service element in a network. The Gamma function was selected by statisticians because of its generality; the Maxwell and Rayleigh distributions were added for completeness.

4.4 STATISTICAL CALCULATIONS

A "zeroth order filter", better known as engineering judgement, was applied to the candidate distributions. The "zeroth order" approach involved calculating the defining parameters of each of the candidate distributions by using the mean and variance of the test data histogram. If the defining parameters were not judged reasonable, i.e. if they were too large or too small, the distribution was discarded. Before using the filter, the 500 SCIS transit times were added and the average found. The variance for these data was also found. These two statistics were used to filter the set of candidate distributions in the following manner.

The calculations for the gamma distribution are shown in figure 17. It can be shown that the gamma distribution has a mean and variance E(T) and var(T) equal to a*b and a*b**2, respectively. The parameters, a and b, are the defining parameters for the gamma function. In the study, the mean found from the data is set equal to a*b and the variance found from the data was set equal to a*b**2. With the values found from the data, the defining parameters, a and b, were calculated to be 50 and .029 respectively. Just as one would not believe that a polynomial of degree 50 is a good fit to the data, these calculated values for the defining parameters indicate that the gamma function probably does not represent the data.

The next candidate distribution to be considered was the Erlang distribution. The calculations for this work are shown in figure 18. The mean and variance for the data were also set equal to the defining parameters of this distribution. The mean, E(T) , was set equal to 1/u and the Var(T) to 1/(k*u**2). The value for u found was .83 and the value for k, 50.05. These values also indicate that the Erlang distribution was not a viable distribution for the SCIS test data.

Calculations for the displaced exponential distribution are shown in figure 19. The approach used in this analysis is slightly different from that taken above. In this approach, three properties of probability density functions are employed. These are 1) that the total area under the curve is equal to 1, 2) that the mean is equal to the integral of t * f(t), and 3) that the variance is the integral of t**2 * f(t) minus the mean squared. Using these well known properties, the values of the defining parameters were found. The parameter, a, the point at which the density function starts, was found to be 1.025. The value of lambda was 5.7 and the value of k, 1993. The parameter, k, seemed too large, and this distribution was placed in a "maybe" category.

In order to determine if the distribution was normal, a range of means and variances was chosen. Since the 99 percent value was known from the specification, the normalization equation, given in figure 20, was set equal to 2.33, the 99.8 percent value of the normalized curve. If the data were normal, the 99.8 percent normalized value derived for the observed

M. H. Weeden

IL2949/ref VL14635

$$E\,(T) = \alpha\beta$$

$$Var\,(T) = \alpha\beta^2$$

$$1.2 = \alpha\beta$$

$$.029 = \alpha\beta^2$$

$$.029 = 1.2\beta$$

$$.024 = \beta$$

$$\frac{1.2}{.024} = \alpha$$

$$50 = \alpha$$

Figure 17. Sample Calculations – Gamma

IL2950/ref VL14394

$$m = \frac{1}{\mu} \qquad \mu = 1.2 \quad \Rightarrow m = .8\overline{3}$$

$$\sigma^2 = \frac{1}{k\mu^2} \qquad k = 50.05$$

Not a reasonable model?

$$f(t) = \frac{(k\mu)^k \, t^{k-1} \, e^{-k\mu t}}{(k-1)!} \qquad \begin{array}{l} t \geq 0 \\ k > 0 \\ \mu > 0 \end{array}$$

Figure 18. Sample Calculations – Erlang

IL2951/ref VL14399

$$f(t) = ke^{-\lambda t} \quad t \geq a$$

$$\int_a^\infty ke^{-\lambda t}\, alt = 1$$

$$\int_a^\infty kte^{-\lambda t}\, alt = 1.2$$

$$\int_a^\infty kt^2e^{-\lambda t}\, alt - m^2 = \sigma^2$$

$a = 1.025$
$\lambda = 5.7$
$k = 1993$

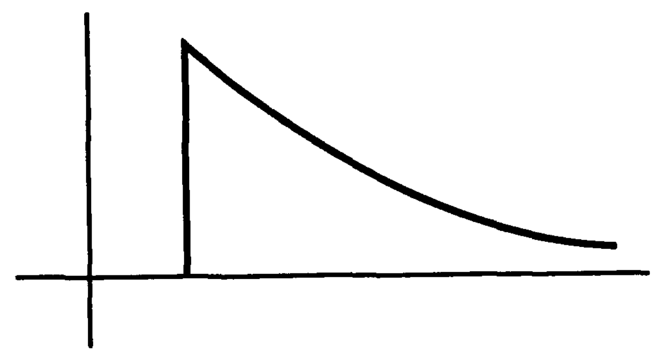


Figure 19. Sample Calculations – Displaced Exponential

IL2952/ref VL14396

$$2.33 \overset{?}{=} \frac{(a-\mu)}{\sigma}$$

μ \ σ^2	S_1^2	S_2^2	S_3^2	S_4^2
$\overline{X}_1$	8.48	6.92	6	5.45
$\overline{X}_2$	7.07	5.77	5	4.54
$\overline{X}_3$	5.66	4.62	4	3.63
$\overline{X}_4$	4.24	3.46	3	2.73
$\overline{X}_5$	3.54	2.88	2.5	2.27

Figure 20. 'Sample Calculations – Normal

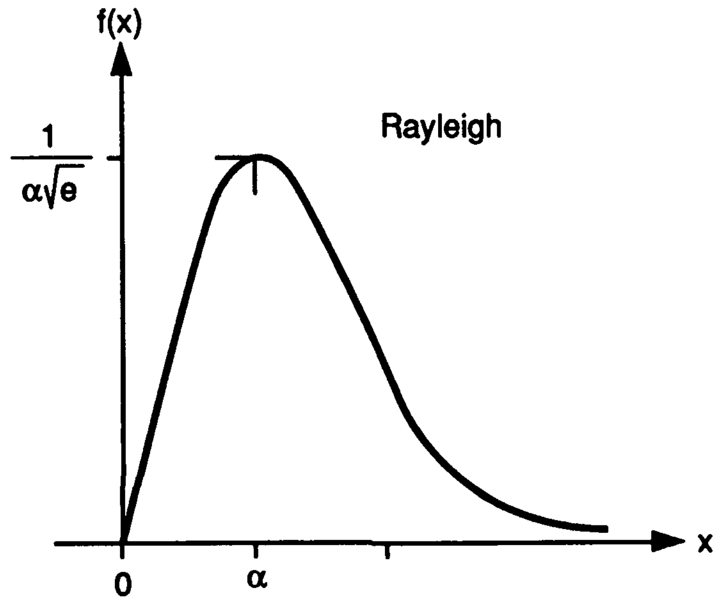
mean and the standard should approximate 2.33. The values found are shown in figure 20. Letters have been substituted for the actual values of the means and variances. One reasonable value, calculated from the normalization equation, was found just outside of the postulated acceptable range of means and variances. The normal curve was placed in a "maybe" category. Further investigation using the Chi Square statistic eliminated this candidate distribution.

In a similar manner, both the Rayleigh and Maxwell distributions were eliminated as shown in figure 21 and 22. In these calculations the mode of each distribution is assumed to occur at the SCIS test data mean transit time. The question was then asked: Does the mode of 124/500, the SCIS test data peak, equal or nearly equal the peak value for the proposed distributions? The answer in both cases is no. These two distributions were eliminated as candidates.

4.5 RESULTS OF JUDGMENTAL FILTER

The zeroth order calculations indicated that no distribution passed the filter. The results from applying the judgmental filter are summarized in figure 23. Since no definitive density function was found, a new approach based on approximations using the specification values was pursued. This approach proved fruitful and is described in the following sections.

IL2953/ref VL14644

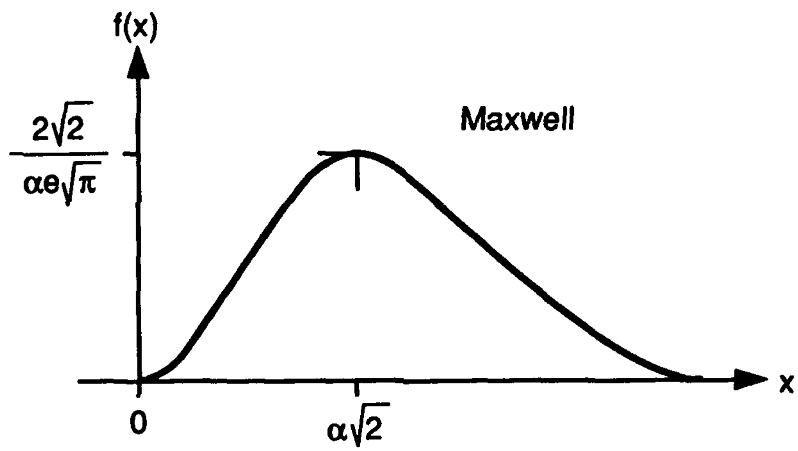


$$f(x) = \frac{x}{\alpha^2} e^{-x^2/2\alpha^2} U(x)$$

$$\text{Does} \quad \frac{124}{500} = \frac{1}{1.2\sqrt{e}} \text{ ? No}$$

Figure 21. Sample Calculations – Rayleigh

IL2954/ref VL14645



$$f(x) = \frac{\sqrt{2}}{\alpha^3\sqrt{\pi}} x^2 e^{-x^2/2\alpha^2} U(x)$$

$$\text{Does} \quad \frac{124}{500} = \frac{\sqrt{2}}{\alpha e\sqrt{\pi}} \ ? \quad \text{No}$$

Figure 22. Sample Calculations – Maxwell

IL2955/ref VL14812

- The "zeroth order" study of the SCIS data indicates that probably none of the candidate distributions "fits" the data

	SCIS			CSSR		
	Yes	No	Maybe	Yes	No	Maybe
Gaussian			✔		✔	
Erlang		✔				✔
Displaced Exponential			✔		✔	
Gamma		✔			✔	
Other		✔			✔	

Figure 23. Zeroth Order Results – Summary

SECTION 5

THE SEARCH FOR APPROXIMATE SOLUTIONS

5.1 THE TRIANGULAR/TRAPEZOID APPROXIMATION

Having found no suitable distribution for the SCIS test data, a search for approximations began. The piecewise linear approximation was developed. In this approach, specification values are used to determine a triangular, or set of triangular, distributions that "encloses" the test data. Instead of trying to find a curve that fit the data, a bound or enclosure of the test data was sought. Then, the test data can vary within the enclosure and have any distribution without affecting the approximate specification distribution.

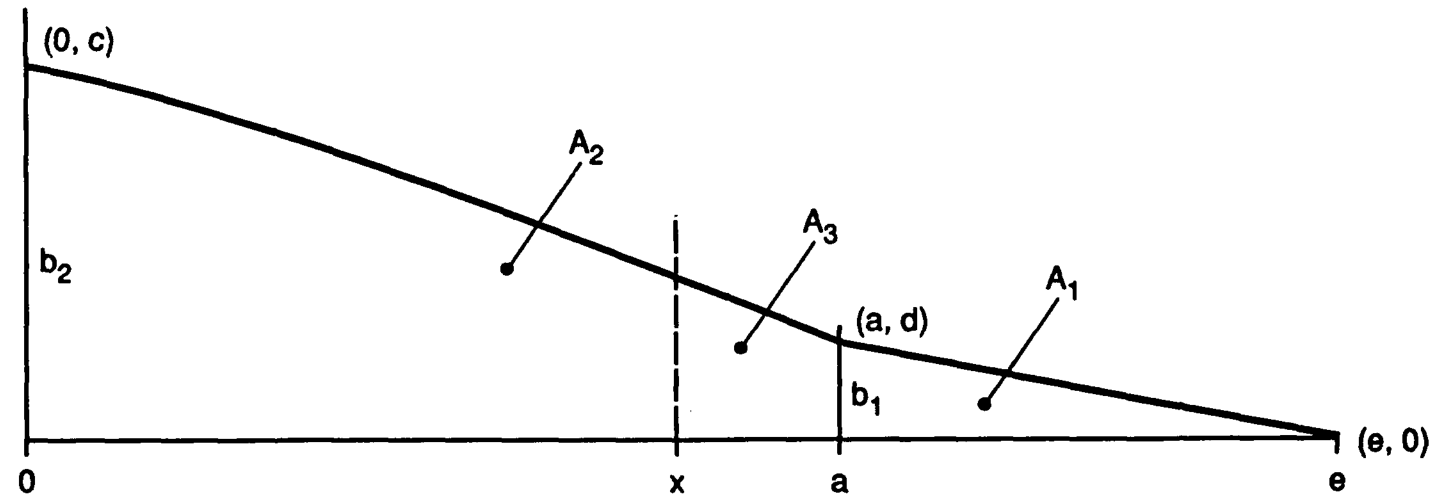
In the triangular distribution, the area of the triangle, (or multiple triangles, or one triangle and one trapezoid) was set equal to the specified percentage. Thus, if a transit time were specified as x seconds in 99 percent of the time, then the area of the triangle was .99 and the end point or vertex of the triangle was the value of the transit time at that specification, x. Using the information about the area and the endpoints, the equation of the lines of each of the triangles was found. From this equation of the line, any percentage value could be found. A sample triangular distribution is given in figure 24.

Even though this approximation technique did not produce a suitable distribution, the ambiguities or lack of information in the subsystem specifications was highlighted. In a number of the subsystems, the length of the tail was unknown. It was impossible to determine the location of the mode or the mean for most of the specifications. This was the same problem encountered in the earlier analysis. The location of the initial or smallest transit time was difficult. The number of modes in the distribution was also unknown. The triangular distribution investigation pointed out how little we know about the distribution. The investigation of the triangular approximation was abandoned and the exploration of the uniform or rectangular distributions began.

5.2 THE RECTANGULAR/UNIFORM DISTRIBUTION APPROXIMATION

Theory indicates that when little is known about the distribution, a uniform distribution, see figure 25, is preferred. In *Introduction to Simulation and SLAM* Pritsker and Pegden [2] say, "The use of the uniform distribution often implies a complete lack of knowledge concerning the random variable other than it is between a minimum and a maximum value." The uniform distribution is a constant value between an beginning value and an end value. This means that if the probability density function is constant over the interval, then the probability of each transit time occurring is equally likely. Moreover, the area under the uniform distribution is one. Since the specifications do not always provide a maximum value for the distribution, the rectangular distribution is used in place of the uniform distribution for some subsystems. The differences between the uniform and the rectangular distribution are 1) the area under a rectangular distribution does not need to be one, and 2) the rectangular distribution can be the concatenation of a number of rectangular distributions as shown in figure 26.

IL2956/ref VL14646



$$A_1 = \frac{1}{2}(e - a)b_1$$

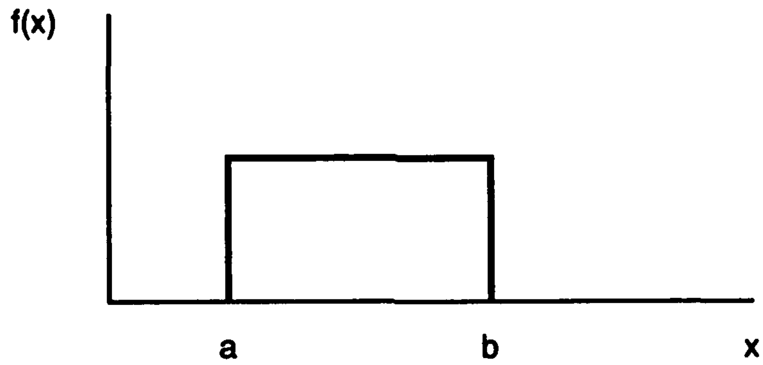
$$A_2 = \frac{1}{2}(b_1 + b_2)a$$

Get equation of lines, then integrate from x to a to get A_3. Solve for x

Shape of triangular tail and trapezoid assumed in some cases

Figure 24. The Piecewise Linear Approximation

IL2957/ref VL14997

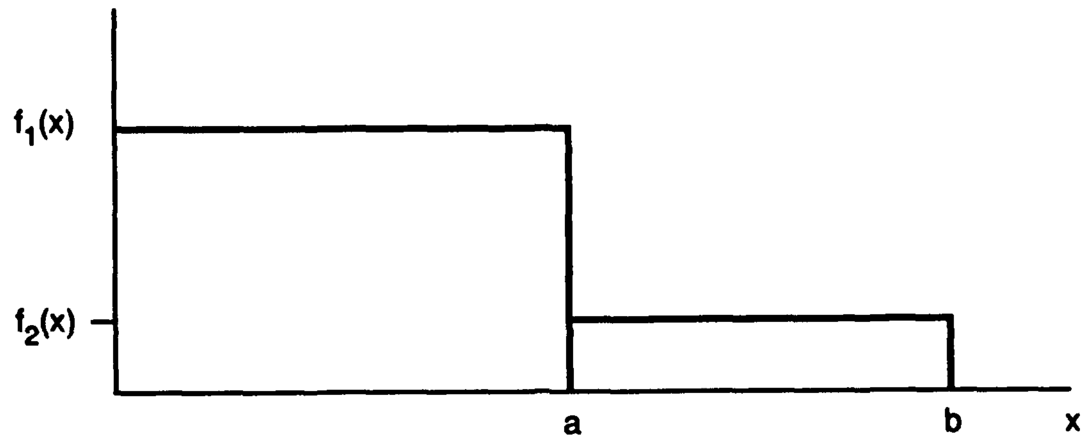


$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{Elsewhere} \end{cases}$$

Figure 25: The Uniform Distribution

M. H. Weeden

IL3035/ref VL15144



$$f(x) = \begin{cases} f_1(x) & 0 \le x \le a \\ f_2(x) & a < x \le b \\ 0 & \text{elsewhere} \end{cases}$$

Figure 26. The Rectangular Distribution

The choice of the rectangular or uniform distribution for the transit time distribution has a number of advantages. Since each of the transit times is equally likely, no bias is introduced by the analyst by making assumptions about the mean or the variance. The uniform (and rectangular) distributions have a minimum nonzero value so that there is zero probability of negative transit times.

5.3 DETERMINATION OF OVERALL INFORMATION DELIVERY TIME

In order to determine the overall information delivery time, the new methodology proposes that the best approximation to the overall delivery time will be found by the convolution of the uniform or rectangular distributions to acquire an overall distribution. This approach is consistent with The Fundamental Theorem on page 189 of *Probability, Random Variables, and Stochastic Processes* by Papoulis [1]. This theorem states: "If the random variables x and y are independent, then the density of their sum z = x + y equals the convolution of their respective densities."

By densities, Papoulis means distributions, or basically histograms. (A histogram is a graph with, in this case, transit times on the independent axis and counts of the occurrences of these times along the dependent axis.) The Fundamental Theorem means that the analyst can not add transit times at a specific percentile from different distributions and report the sum as either a worst case or the corresponding percentile. There are two exceptions. It can be shown that the means and, in the case of independence, variances can be added to obtain the mean and variance of the overall distribution.

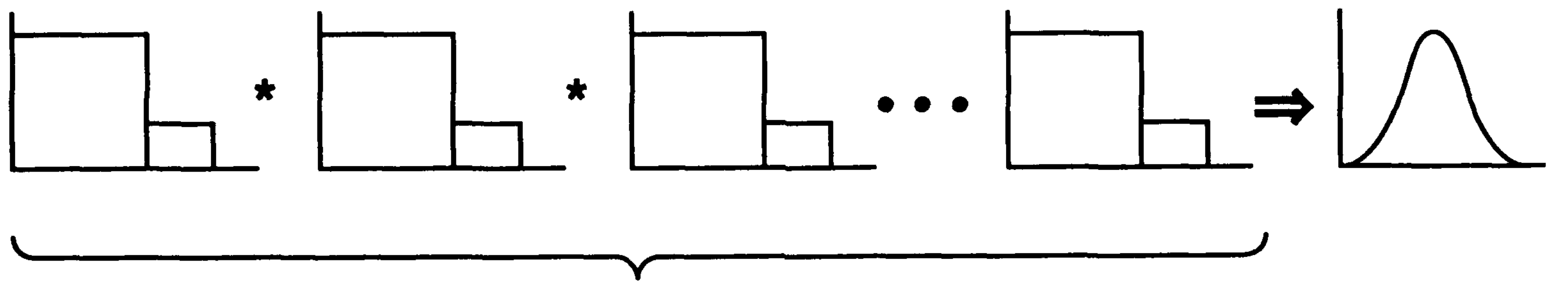
With these theorems in mind, the overall information delivery time density function can be found by convolving the transit time distributions for each subsystem as shown in figure 27. The number of convolutions and the distributions involved in those convolutions are dependent upon what subsystems are in each string. In the case of the missile warning string, there are eight traversals of the various subsystems; for air and space, there are six. This means that the resultant distribution is the convolution of either six or eight distributions. Since convolution is difficult, especially for a convolution of six or eight distributions, an approximation to the overall distribution using the Central Limit Theorem can be applied, if an assumption of independence of each subsystem is made, and if the variances of each of the distributions are similar.

The Central Limit Theorem says,

> Let x1, x2, x3, ..., xn be independent random variables. Then whatever the form of their distribution-subject to certain very general conditions-the sum of xn approaches a normal distribution as n increases without bound. This normal distribution has mean equal to the sum of the means and variance equal to the sum of the variances of the n random variables. (*Mathematics Dictionary*, James and James, 1968, p.44) [3]

M. H. Weeden

IL2958/ref VL14813



8 Times for MW
6 Times for AW
6 Times for SW

- Convolve rough approximations to the specification values to obtain an overall specification based information delivery time
- Either tails have to be estimated or ignored in 4 cases
- How quickly does convolution product approach normal, especially in the tails?

Figure 27. Overall Information Delivery Time Density Function

Papoulis shows that convolution of a number of distributions, including uniform distributions, quickly produces a curve that approximates the normal curve. (Significant differences may occur in the tails in some cases, especially if the one of the curves in the convolution has a large variance relative to the other variances. If this situation occurs, as it did in our analysis, then convolution of the transit time distribution with the large variance should be convolved with the Central Limit Theorem approximation to the convolution of the remaining distributions.) Papoulis performs a sample convolution of the uniform distribution with itself on pages 267 and 268 of *Probability, Random Variables, and Stochastic Processes*. [1] This convolution is shown in figure 28. The first convolution produces a triangular distribution. Of the results of the uniform with the triangular distribution, Papoulis says, "...after just two convolutions, the resultant curve is remarkably close to the normal curve."

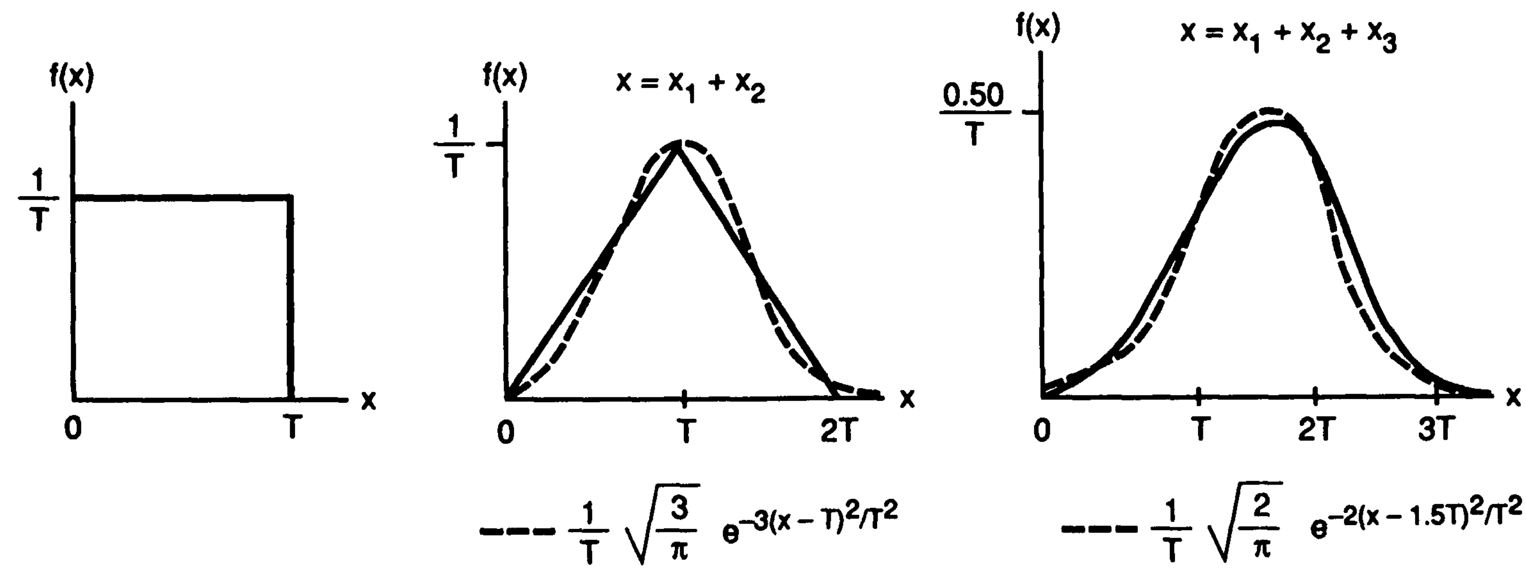
To use the Central Limit Theorem, the mean and variance of each subsystem's uniform distribution must be found. This is done by using the transit time provided in the system specification. If, for example, the maximum transit time for a hypothetical subsystem were 100 seconds and the minimum 0 seconds, then the probability density function for this distribution would be a constant, uniform distribution of 1/100. This constant function, multiplied by x, is then integrated from zero to the end point of the uniform distribution to determine the distribution's mean. To find the variance, the constant probability density function is multiplied by x**2 and integrated over the same interval. The mean squared is subtracted from the result. These are standard methods for finding the mean and variance of a distribution and are depicted in figure 29.

After the mean and variance for each subsystem's uniform, or rectangular distribution, are determined, they are added up over the missile, air and space strings to acquire the mean and variance for the overall message delivery time for those strings respectively. It is correct to add the means of probability density functions. (*Probability, Random Variables, and Stochastic Processes* p. 143 of Papoulis [1]). It is correct to add the variances of probability density functions, if the subsystems are independent. (*Probability, Random Variables, and Stochastic Processes* p. 211, Papoulis [1]). Since the convolution product is nearly normal and since the normal curve is completely determined by its mean and variance, the 99.8 percent value for the overall message delivery time can be found by adding 2.88 times the standard deviation to the mean. The value, 2.88, is the 99.8 percent tabled value for the normalized Gaussian. Using this methodology, any other percentile value can also be found for the overall message delivery times.

5.4 THE EFFECT OF LARGE VARIANCE

In case of the space warning system, a large variance in the uniform distribution caused an inaccuracy in the tails of the overall message delivery normal curve when using the Central Limit Theorem approximation to the overall information delivery time. When the message delivery time was found by convolving the SPADOC subsystem's uniform distribution with the approximation to the normal curve found from the other subsystem's distributions using the Central Limit Theorem approach, a value smaller than the current estimate occurred. This convolution product is more accurate and is the preferred answer in this case.

IL2959/ref VL14726



- Chi-square calculations indicate f(x) and normal curve not significantly different (99%)

Figure 28. Sample Convolution

M. H. Weeden

IL2960/ref VL14985

- Expected value

$$E\{x\} = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

where f(x) is probability density function of x

p. 138 Papoulis

- Variance

$$Var\{x\} = \int_{-\infty}^{\infty} x^2 f(x)\, dx - [E\{x\}]^2$$

p. 144 Papoulis

Figure 29. Standard Methods for Calculating Mean and Variance

5.5 THE SEPARATION OF SPECIFICATION DETERMINATION FROM PERFORMANCE ESTIMATION

Determination of the specification values for the APB values and the quarterly estimates should be separated. This separation is shown graphically in figure 30. The overall message delivery time specification for the APB should be found by the convolution (or Central Limit Theorem approach) of uniform distributions as described above. The distributions for this work should be uniform, or rectangular, distributions drawn from the subsystem specifications. To find the quarterly estimates, real test data, or the means and variances of the real test data, should be used in conjunction with the Central Limit Theorem approach. For those programs that can not supply either real data or the means and variance from real data, the specification uniform distribution can be substituted until the real data becomes available.

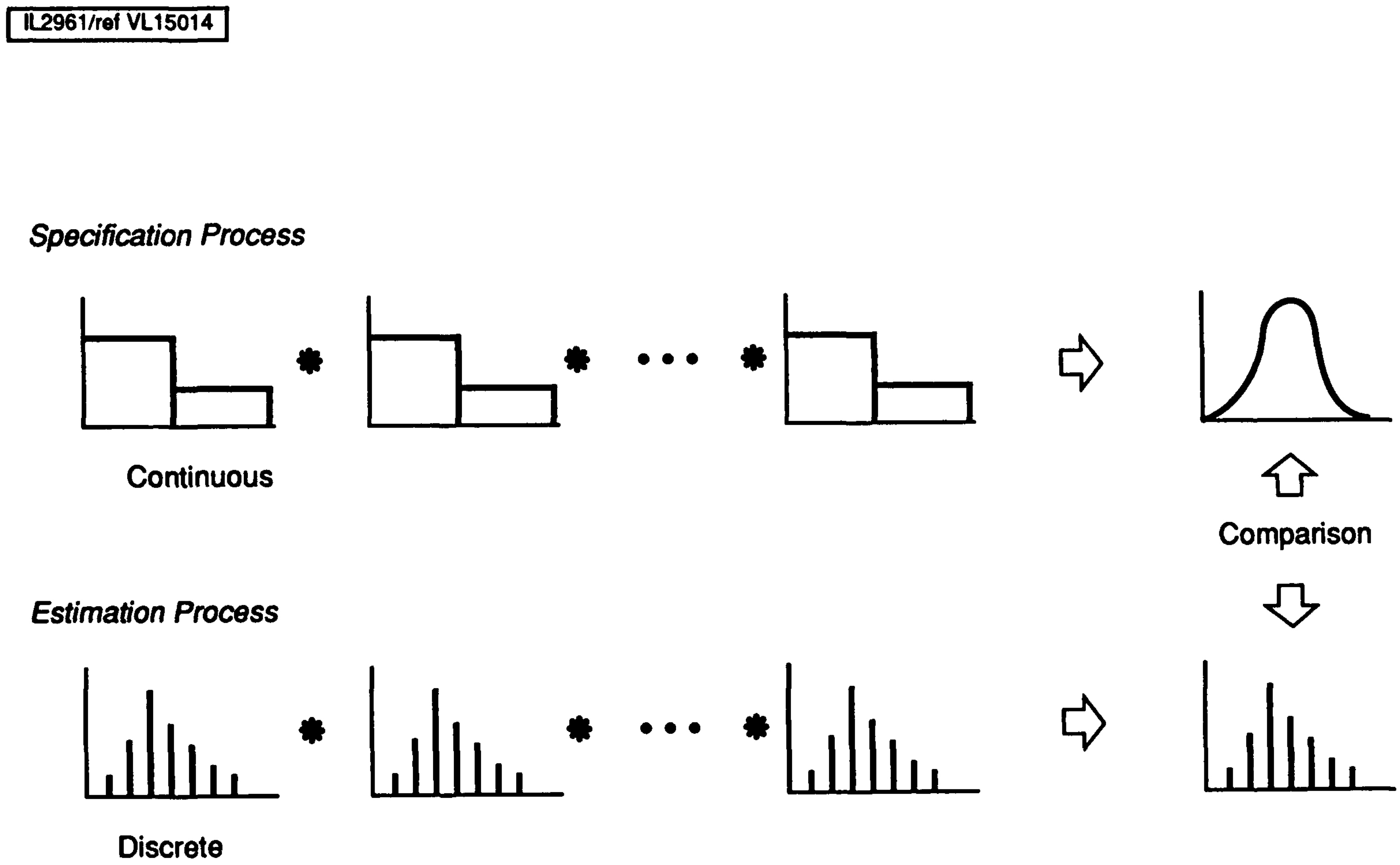


Figure 30. Separation of Specification Determination from Performance Estimation

SECTION 6

CONCLUSIONS

6.1 COMPARISON OF THE METHODOLOGIES

The above methodology differs from the current methodology. A comparison of 99.8 percent values shows that 1) in the missile warning case, a -11 percent difference occurs; 2) in the air warning system, a -9.1 percent change is found; and 3) in the space warning system, a -2.3 percent change is determined. The new methodology produces a smaller overall message delivery time for all three strings. The average decrease was -7.5 percent. See table 1.

6.2 STRENGTHS AND WEAKNESSES OF THE UPDATED METHODOLOGY

The use of the new methodology invoking the convolution of uniform or rectangular distributions, and consequently the Central Limit Theorem, has enhanced statistical rigor. If the means and variances of the test data are supplied by the contractors, the quarterly estimates are simple and easy to do.

Certain assumptions are associated with this methodology. These are 1) that the subsystems are independent and 2) that no subsystem has a distribution with a long tail. Long tails would arise in situations were there is "clogging" in the system. The lack of independence of subsystems could arise if the transit time of a message through one subsystem affected its transit time through the next subsystem. Thus, some sort of history of what had transpired would negate the independence assumption.

The reader should note that the uniform, or rectangular, distribution approach is an approximation to the overall information delivery time distribution. As more information becomes available, improvements to the methodology should be considered.

6.3 RECOMMENDATIONS

The Central Limit Theorem approach to approximating the overall information delivery time was presented to several audiences. It was generally recognized that the new approach possessed enhanced statistical rigor. The approach was recommended as a replacement for the existing methodology.

IL2977

Table 1. Technique Comparison at 99.8%

	Original	*Update*	*% Change*
MW	x	y	–11%
SW	z	w	–2.3%
AW	q	p	–9.1%

$$\% \text{ Change} = \frac{\text{Original} - \text{Update}}{\text{Original}}$$

Average Change –7.5%

LIST OF REFERENCES

1. Papoulis, A., 1965, *Probability, Random Variables, and Stochastic Processes,* New York: McGraw-Hill Book Company.

2. Pritsker, A. B., Pegden, C. D., 1979, *Introduction to Simulation and SLAM*, New York: John Wiley & Sons 1979.

3. James, G., James, R. C., 1968, *Mathematics Dictionary*, Van Nostrand Reinhold Company.

SECTION 8

GLOSSARY

APB	Acquisition Program Baseline
AW	Air Warning
CMU	Cheyenne Mountain Upgrade
CSSR	Communication System Segment Replacement
DCP	Decision Coordinating Paper
ITW/AA	Integrated Tactical Warning/Attack Assessment
MW	Missile Warning
PMD	Program Management Directive
SCIS	Survivable Communications Integration System
SORD	System Operational Requirement Document
SPADOC	Space Defense Operations Center
SW	Space Warning
TEMP	Test and Evaluation Master Plan